DeepLearning Lab3 Report

Chien-Hsun Lai (0656078)

Introduction

In this lab, we use attention model to generate image captions. Give an image as input. The network can produce a sentence describing the input image. With attention mechanism, the network is able to focus on the essential part of the image. The attention mechanism is done by using a 0~1 mask on the input feature. Ideally if some features are important, the network will give it weights close to 1. Moreover, we can visualize what the network is focusing during the caption generation. So that we can check why the network output certain sentences, find out what was wrong.

In this lab, I only experiment the “show attend and tell”

Experiment setup

I uses this code to run the experiment
https://github.com/ruotianluo/ImageCaptioning.pytorch
Only a small parts of the code are modified, I logged the attention manually using bilinear UpSampling to input image’s size then apply Gaussian blur to the attention weight for more visual appealing visualization.

Detail of the model

First, the input image will feed into a ResNet-101 outputs 14*14*512 feature maps. Then this feature map is masked with attention weights using attention mechanism. The weights are computed by a fully connected network Activated with Softmax. After the feature map is masked, it is fed into the decoder LSTM. Finally at the output end. We use beam search to generate the output sequence.

Hyper-parameters

RNN hidden size: 512
RNN layer: 1
RNN type:lstm
Word embedding size: 512
Attention hidden size: 512
Batch size: 16

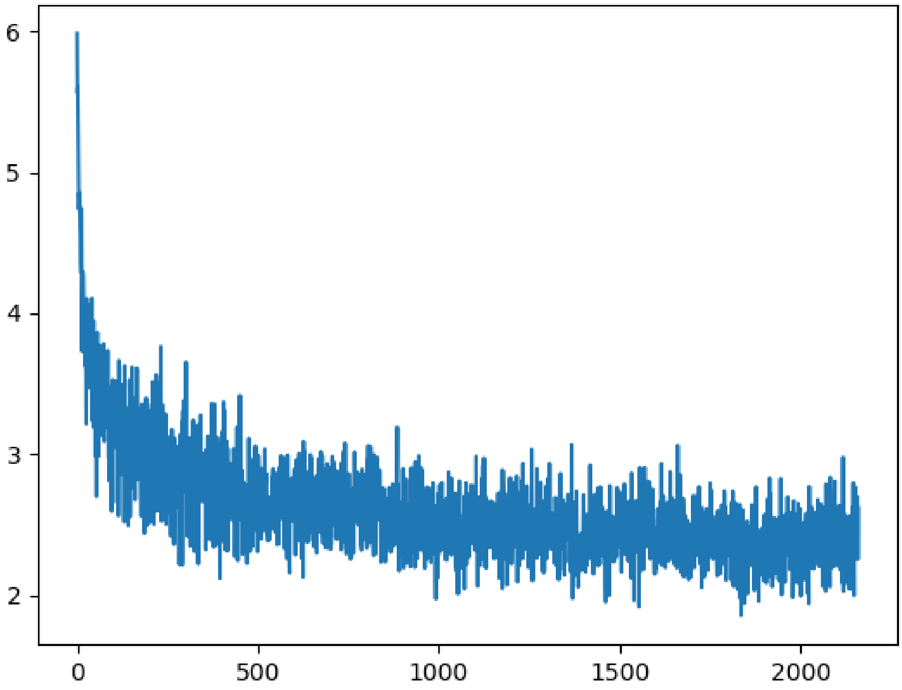
Epoch: 5

Dropout rate: 0.5

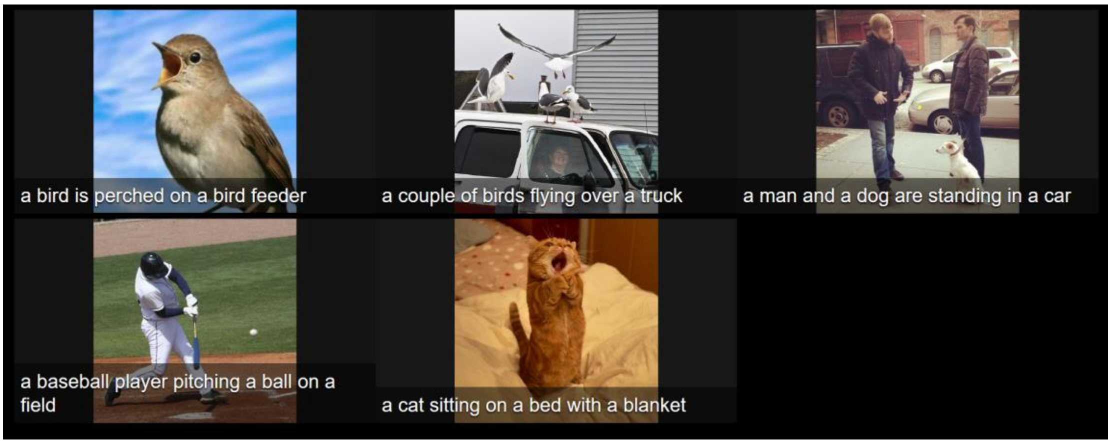
Optimizer: Adam (lr=4e-4)

Result

Training loss of attention model



Caption of model



a bird is perched on a bird feeder

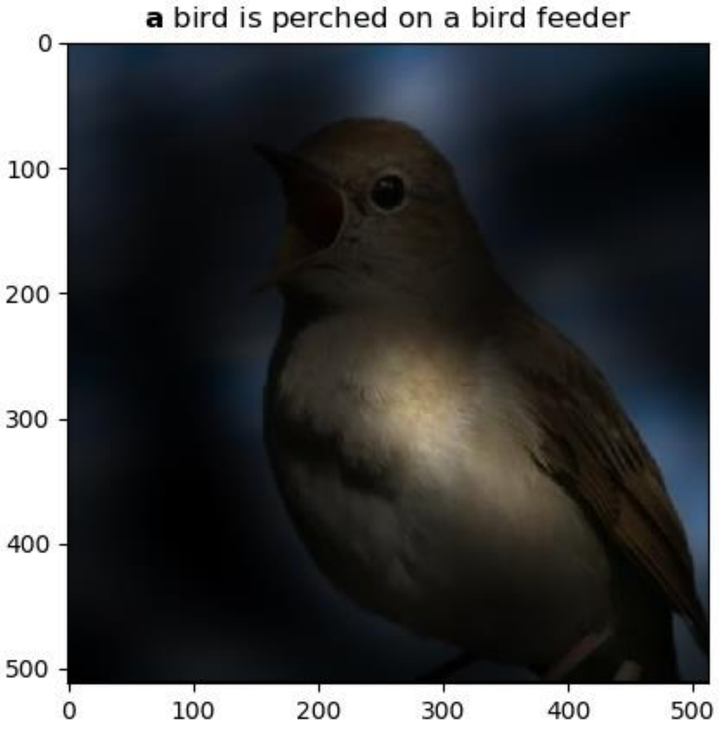
a couple of birds flying over a truck

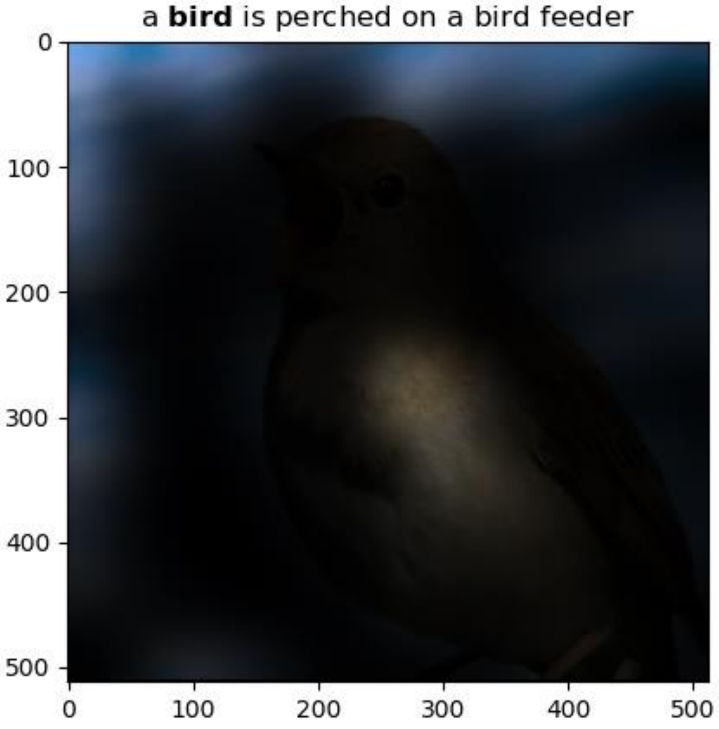
a man and a dog are standing in a car

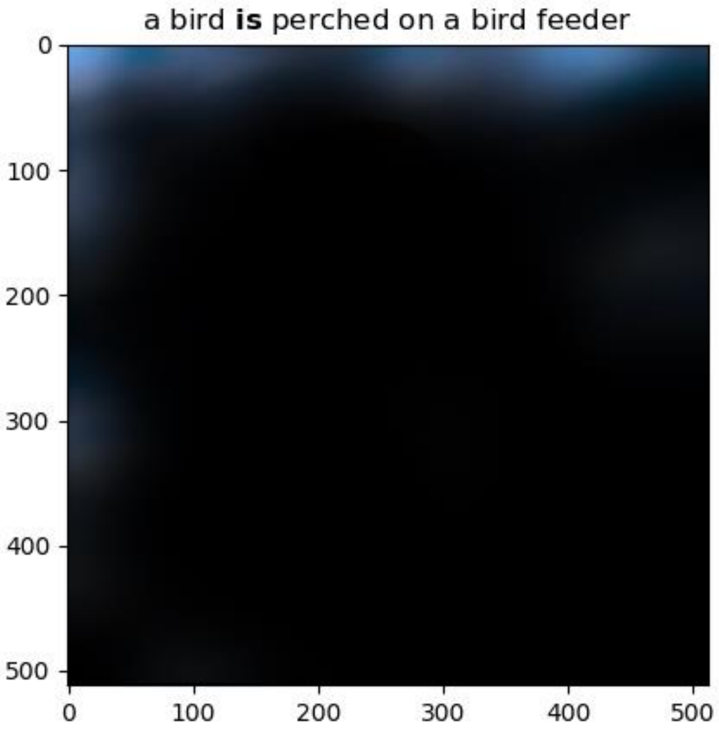
a baseball player pitching a ball on a field

a cat sitting on a bed with a blanket

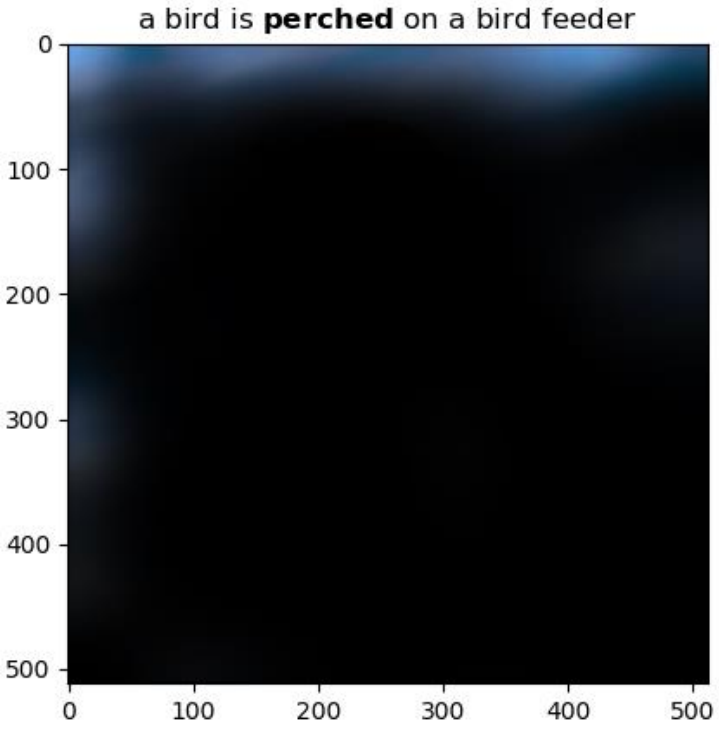
Attention over time



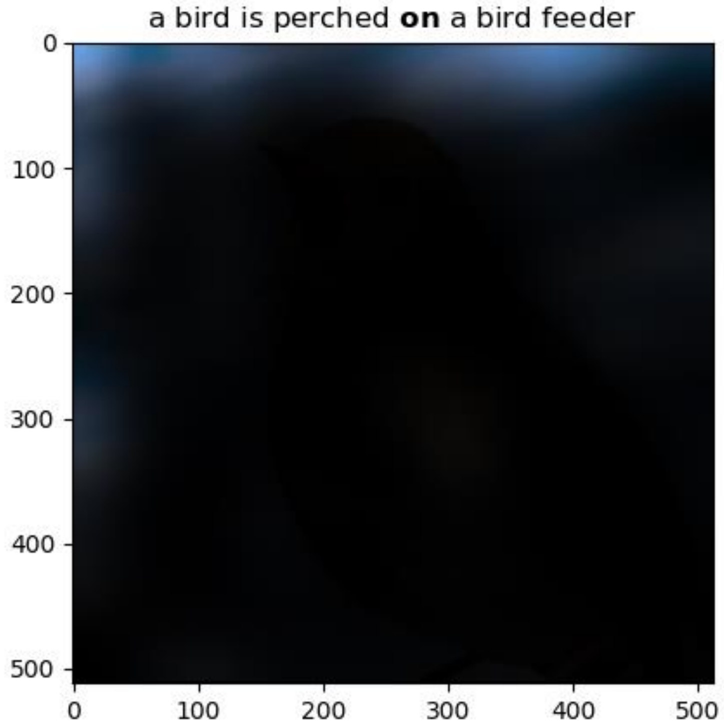


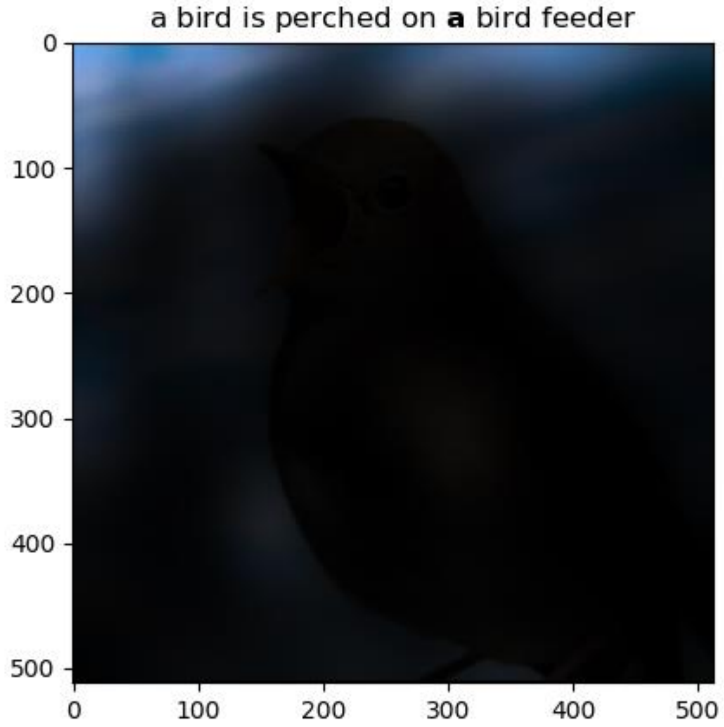


a bird **is** perched on a bird feeder

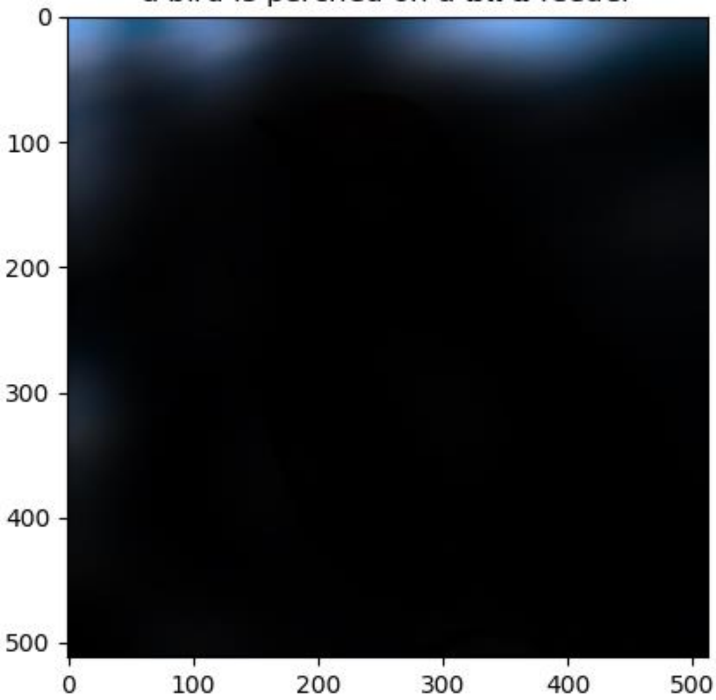


a bird is **perched** on a bird feeder

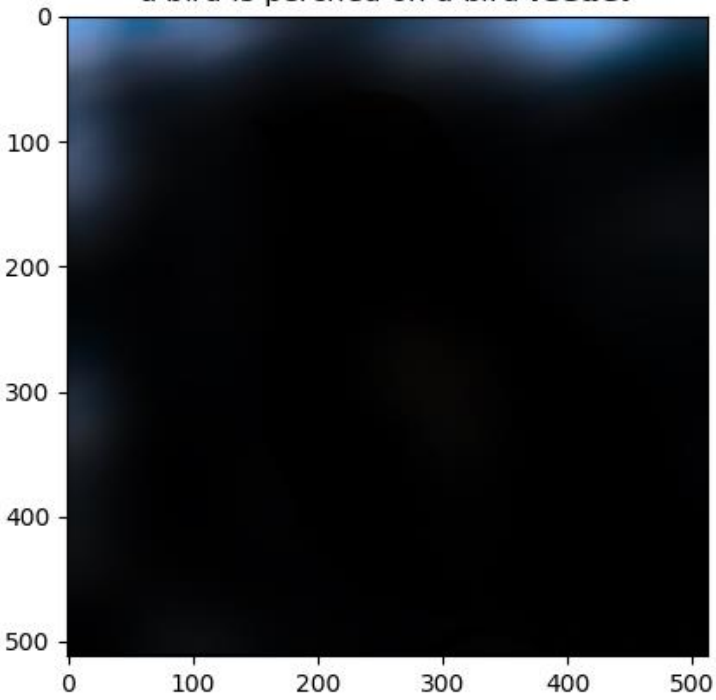

a bird is perched **on** a bird feeder

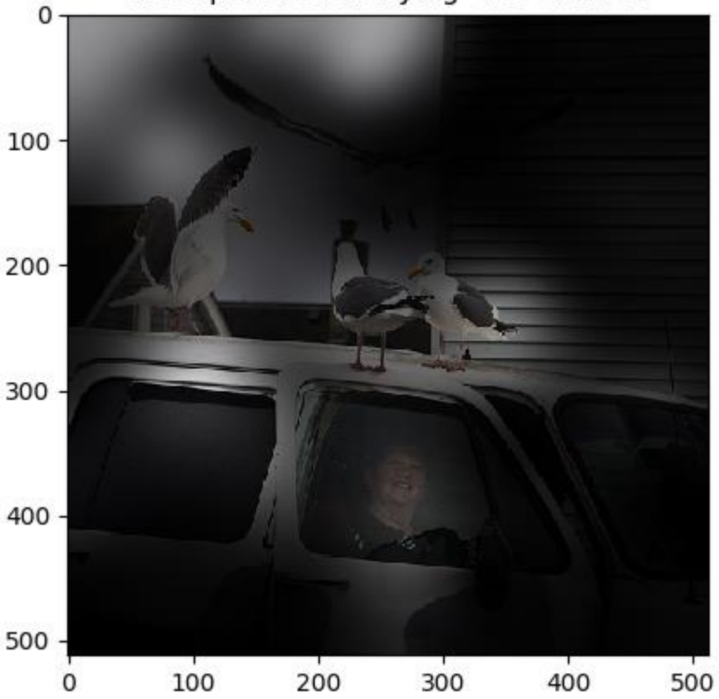

a bird is perched on **a** bird feeder

a bird is perched on a **bird** feeder

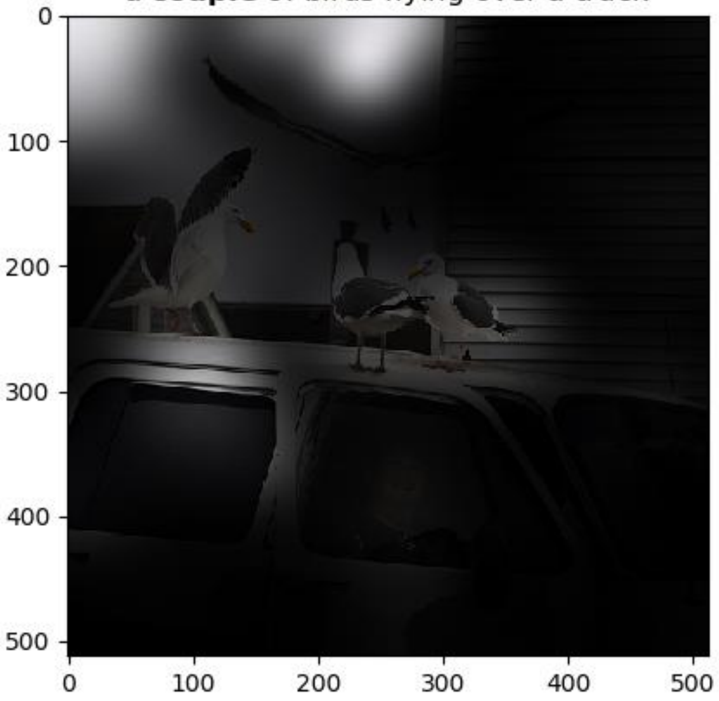


a bird is perched on a bird **feeder**

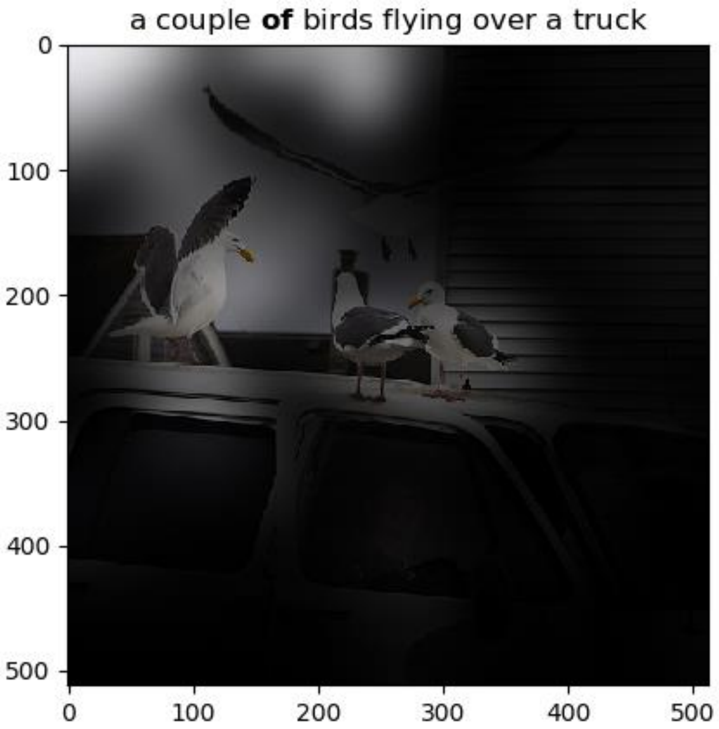


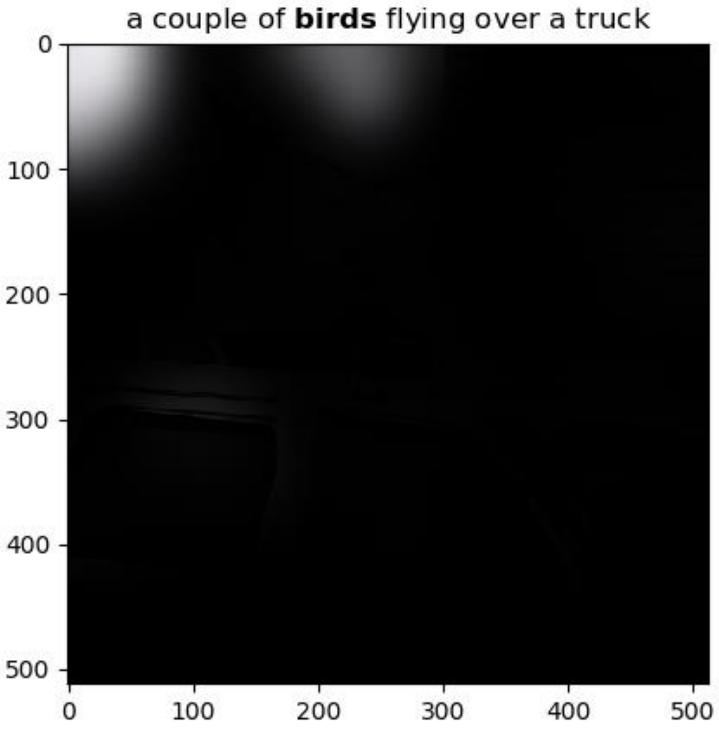


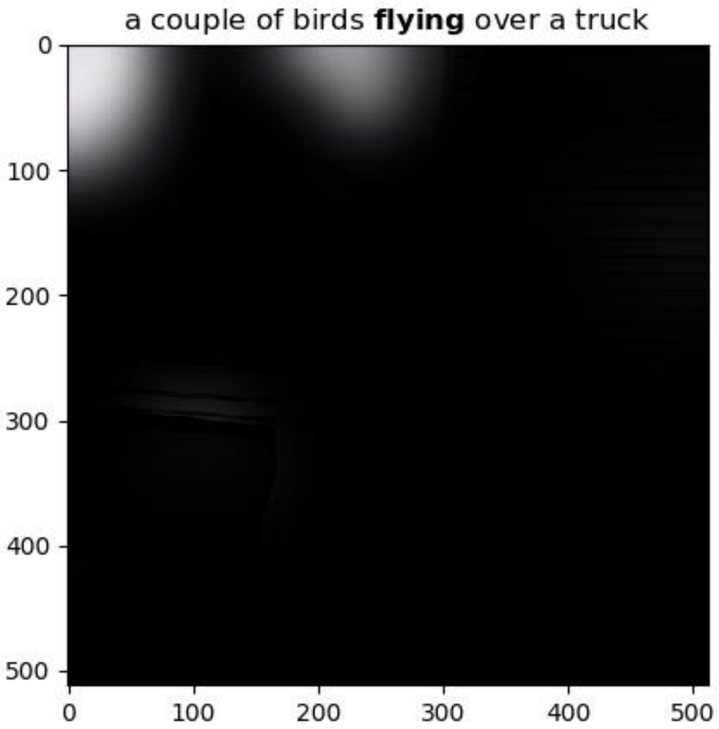
a couple of birds flying over a truck



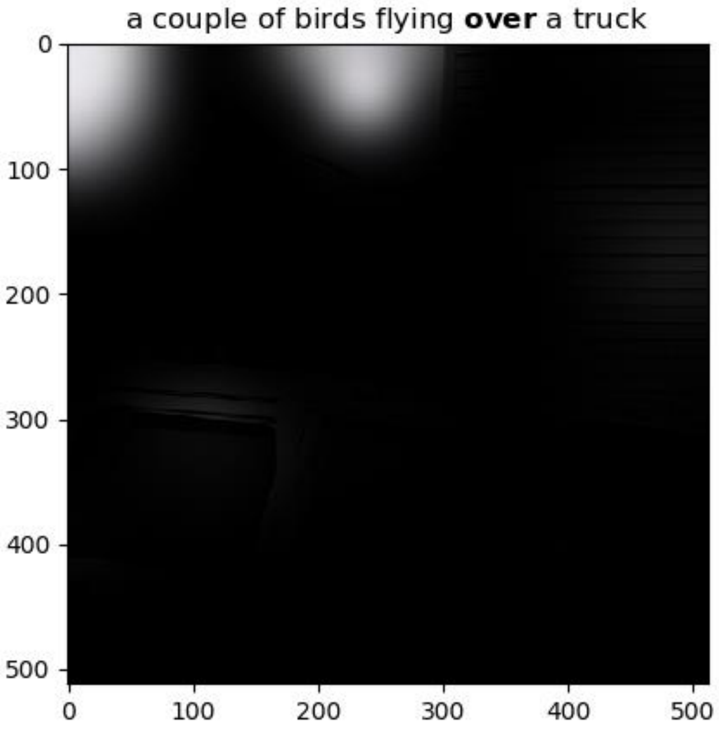
a **couple** of birds flying over a truck


a couple **of** birds flying over a truck
0
100
200
300
400
500
0
100
200
300
400
500

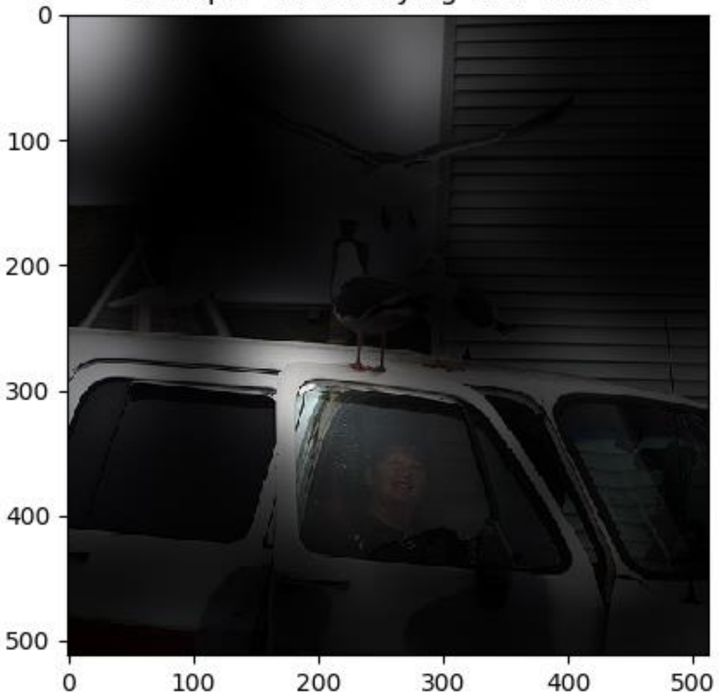

a couple of **birds** flying over a truck
0
100
200
300
400
500
0
100
200
300
400
500



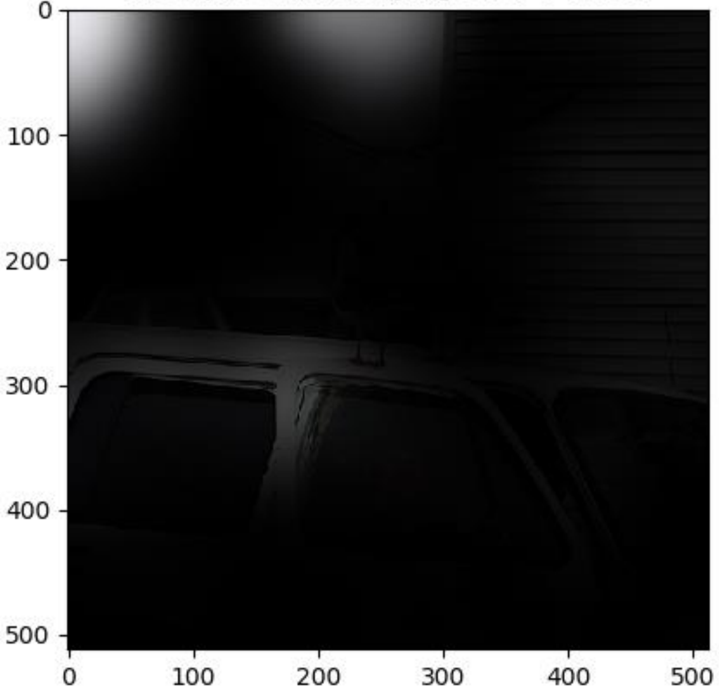
a couple of birds **flying** over a truck

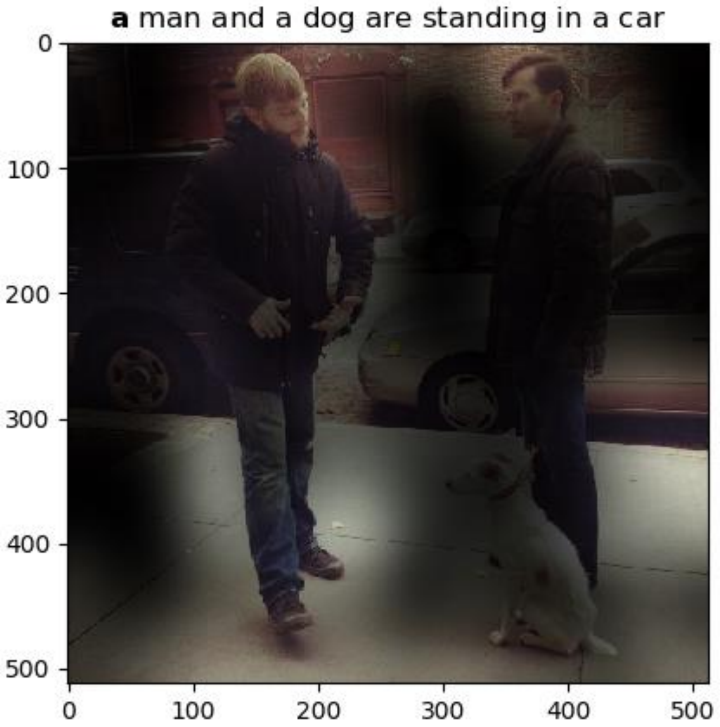


a couple of birds flying **over** a truck a couple of birds flying over **a** truck

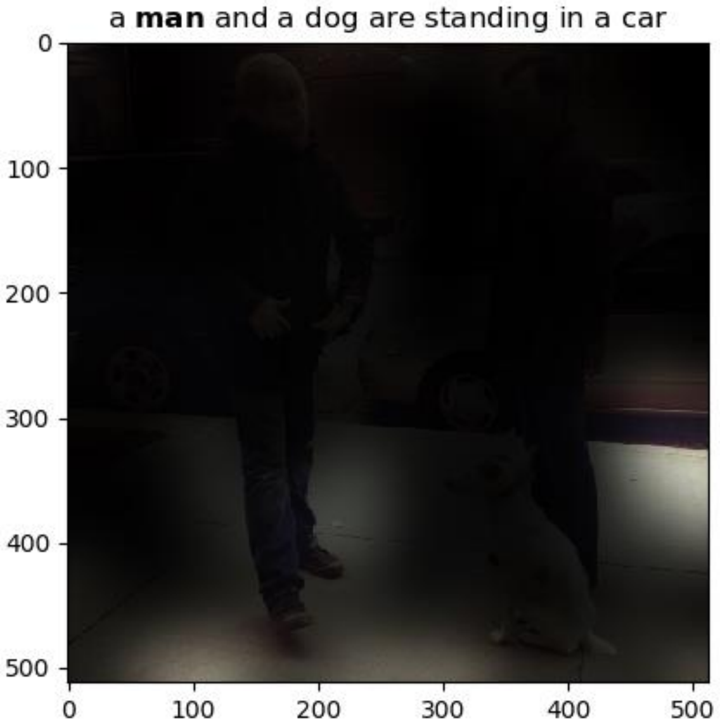


a couple of birds flying over a **truck**

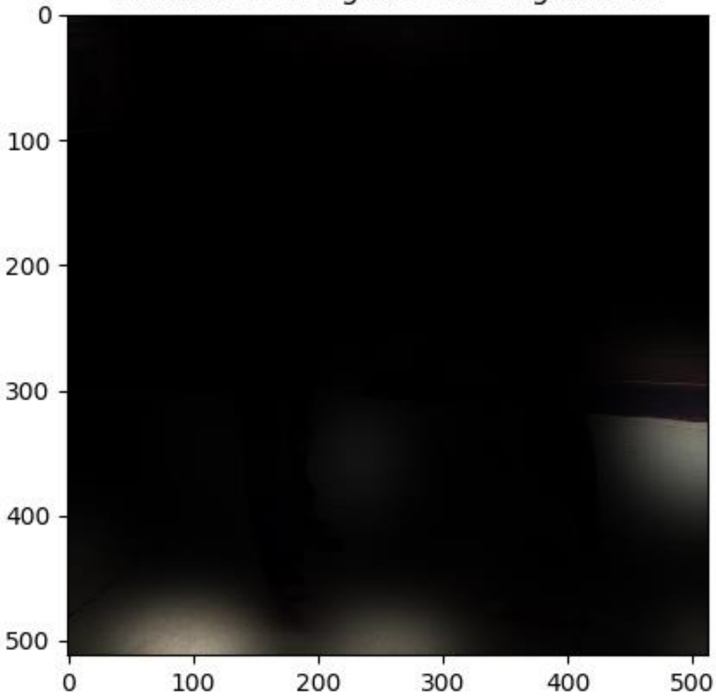




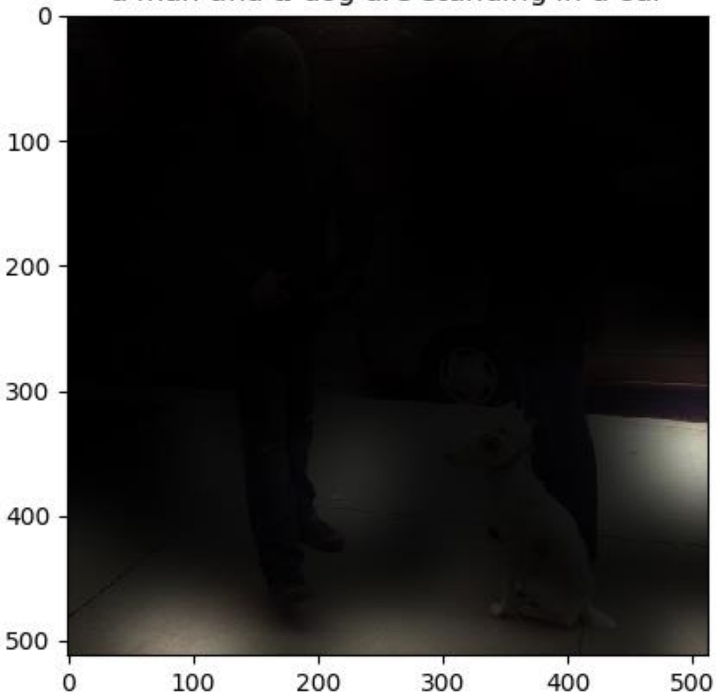
a man and a dog are standing in a car



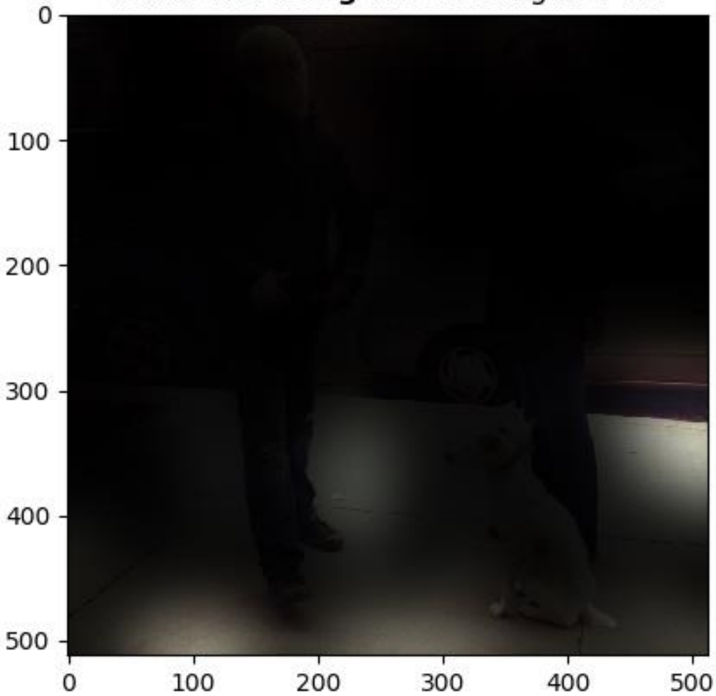
a **man** and a dog are standing in a car



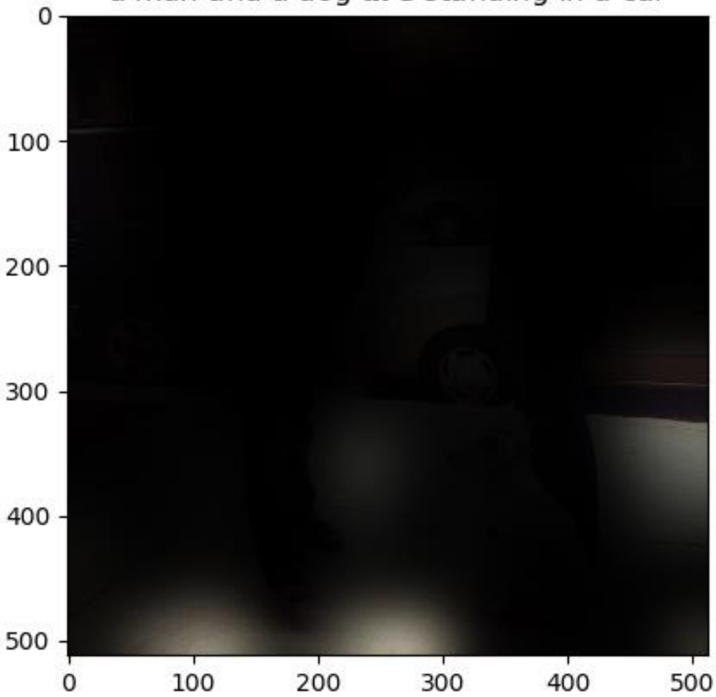
a man **and** a dog are standing in a car

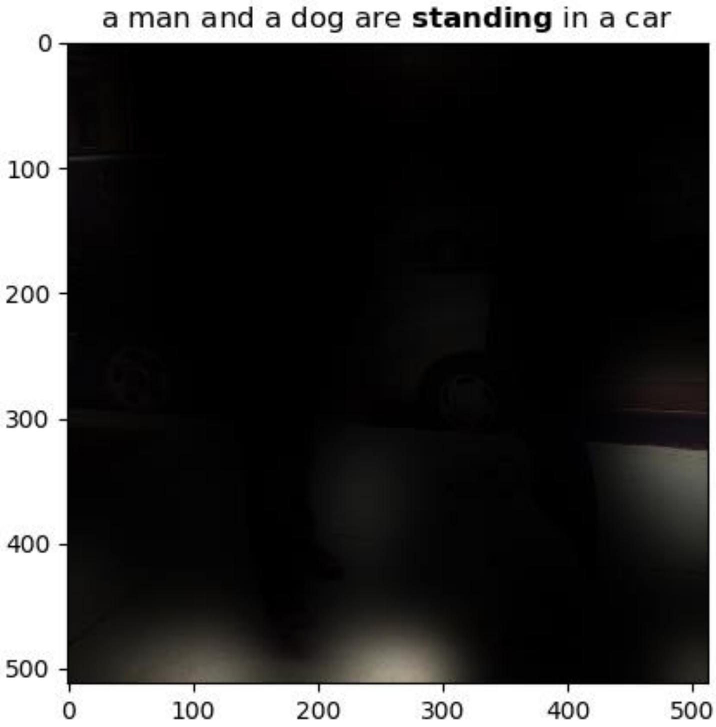


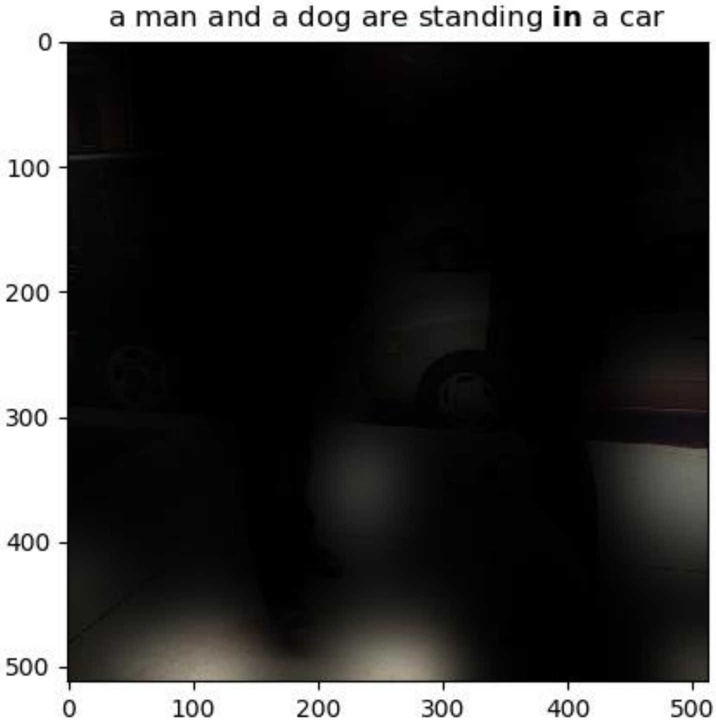
a man and **a** dog are standing in a car a man and a **dog** are standing in a car

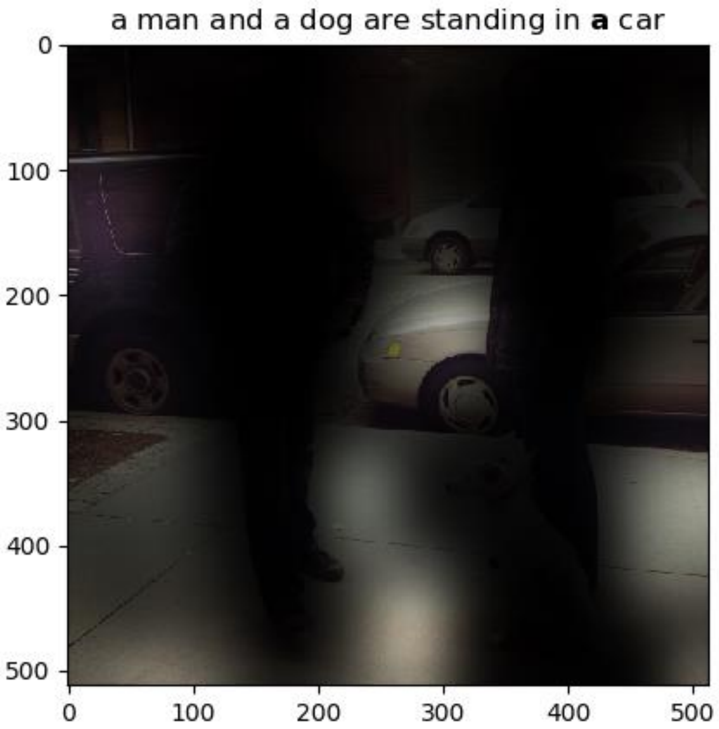


a man and a dog **are** standing in a car

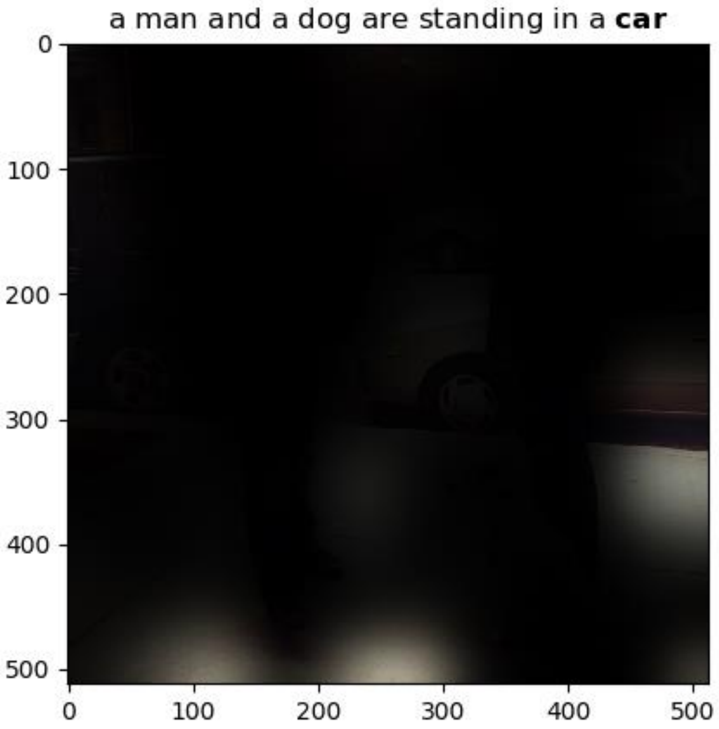



a man and a dog are **standing** in a car

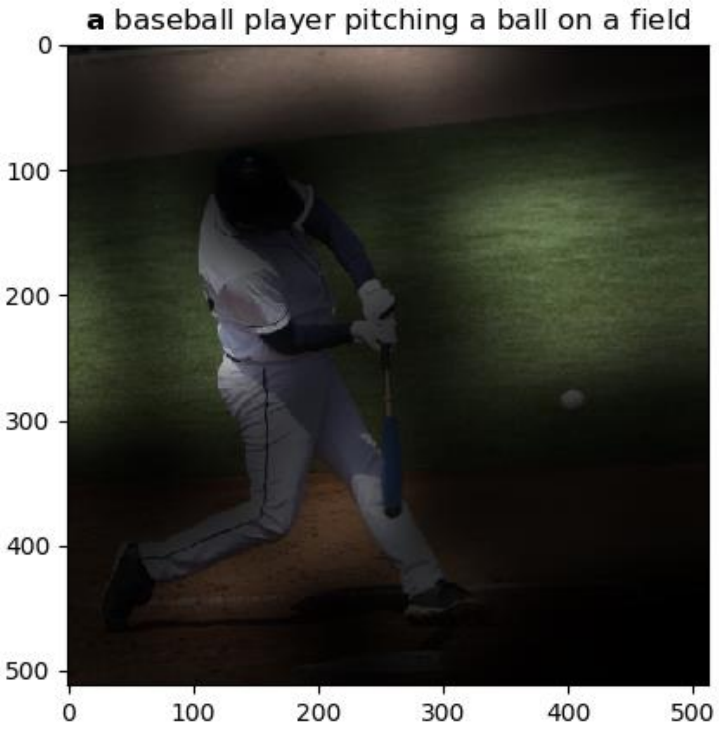

a man and a dog are standing **in** a car



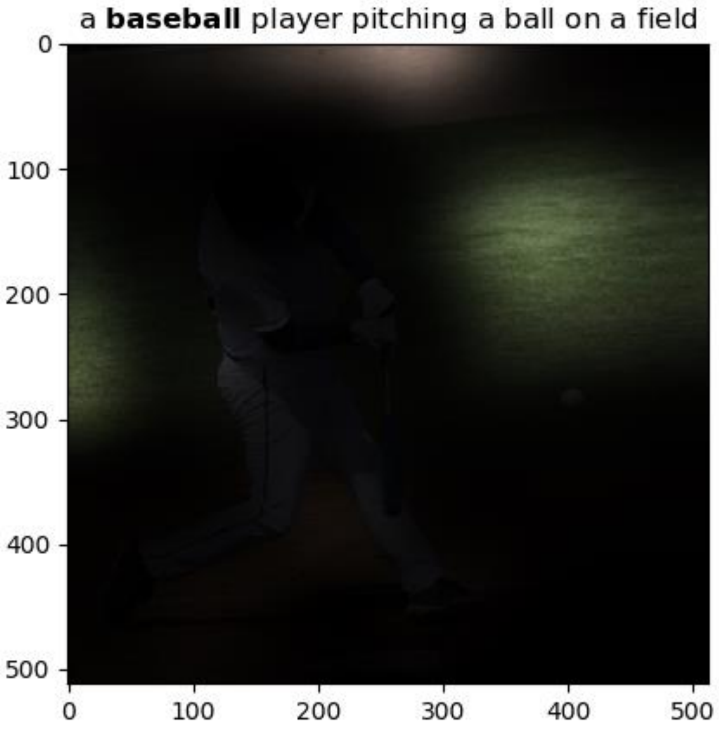
a man and a dog are standing in **a** car



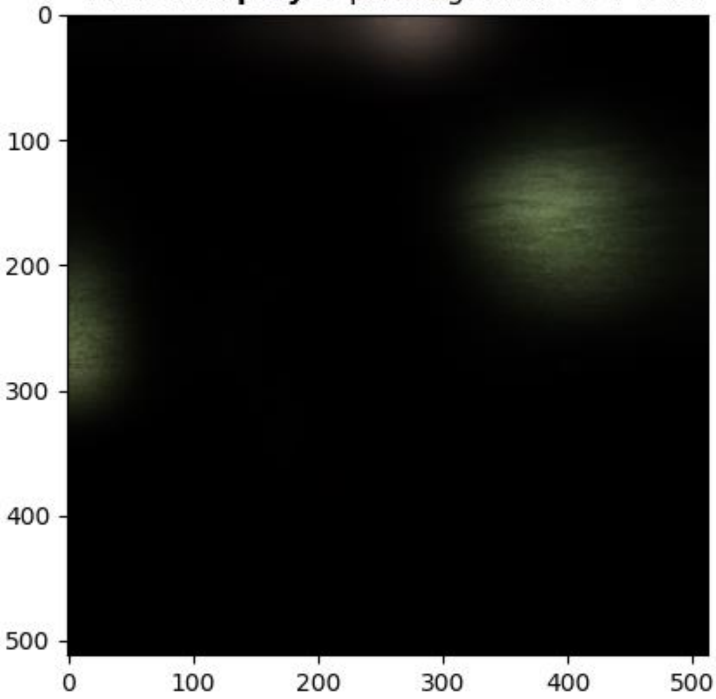
a man and a dog are standing in a **car**



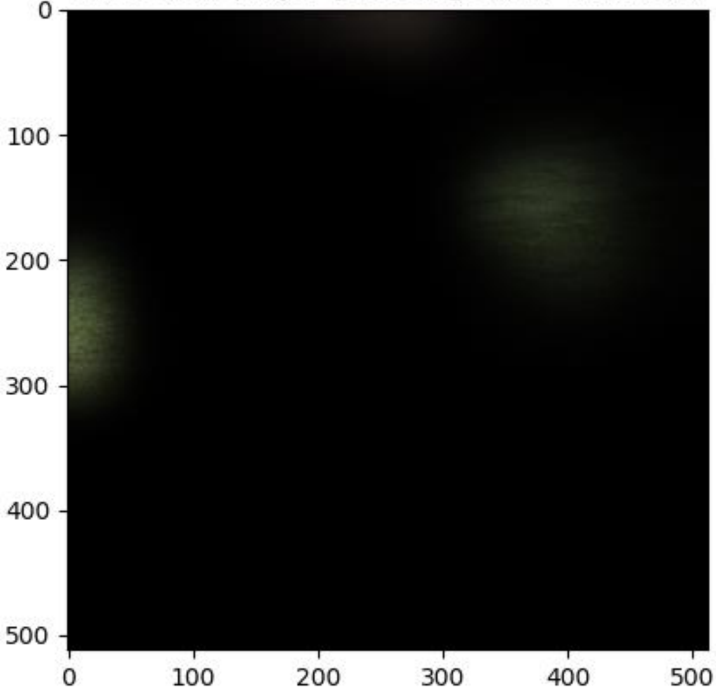
a baseball player pitching a ball on a field



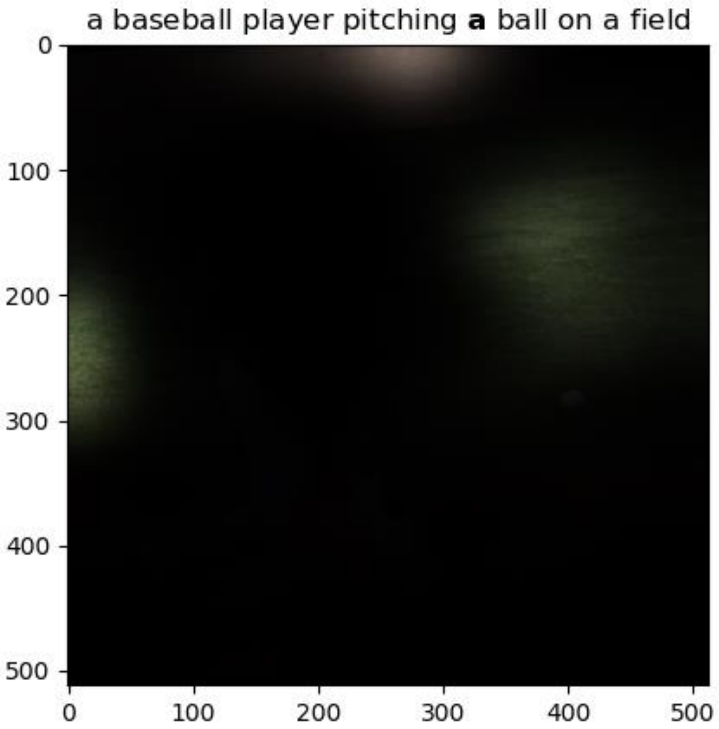
a **baseball** player pitching a ball on a field



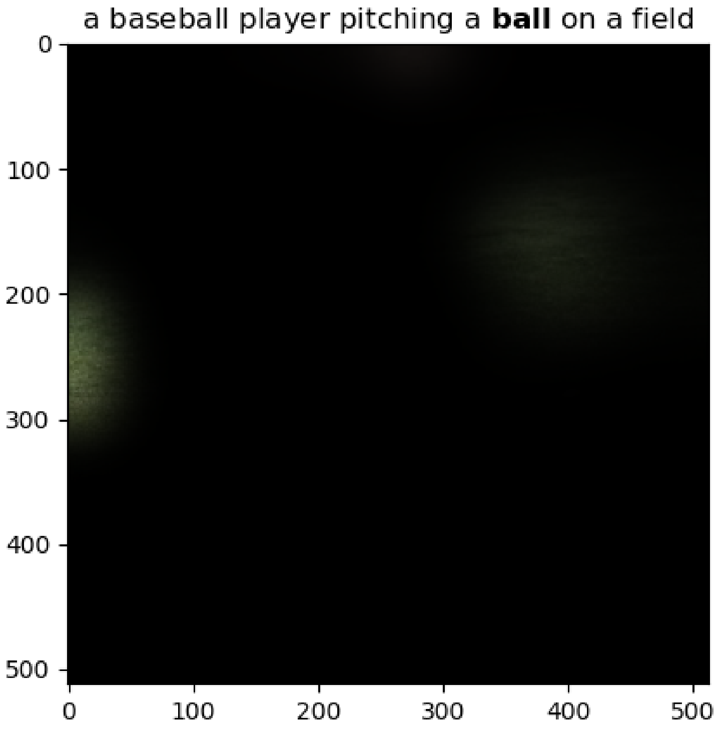
a baseball **player** pitching a ball on a field



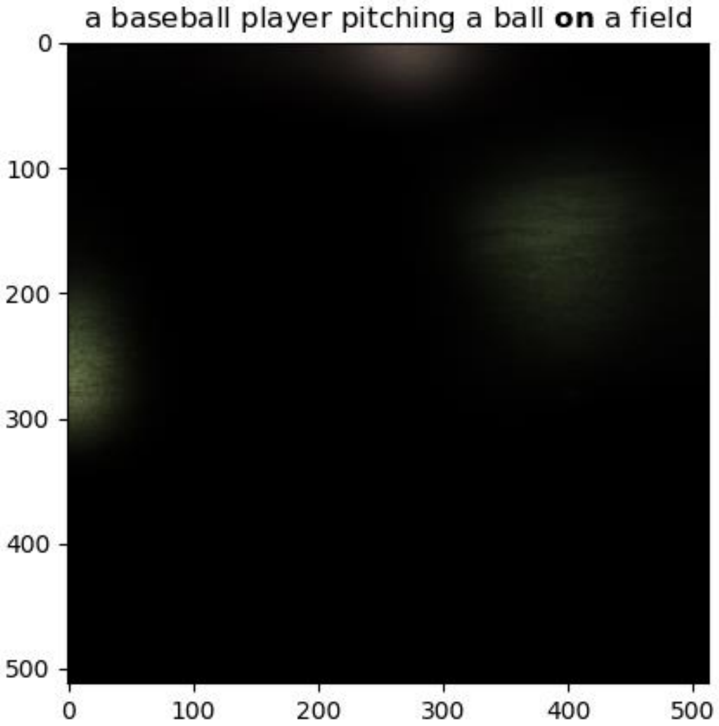
a baseball player **pitching** a ball on a field

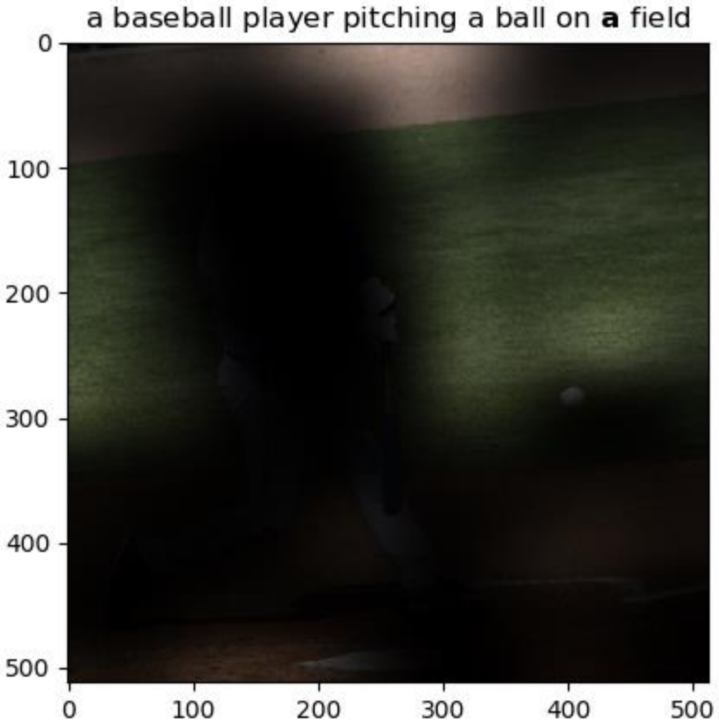


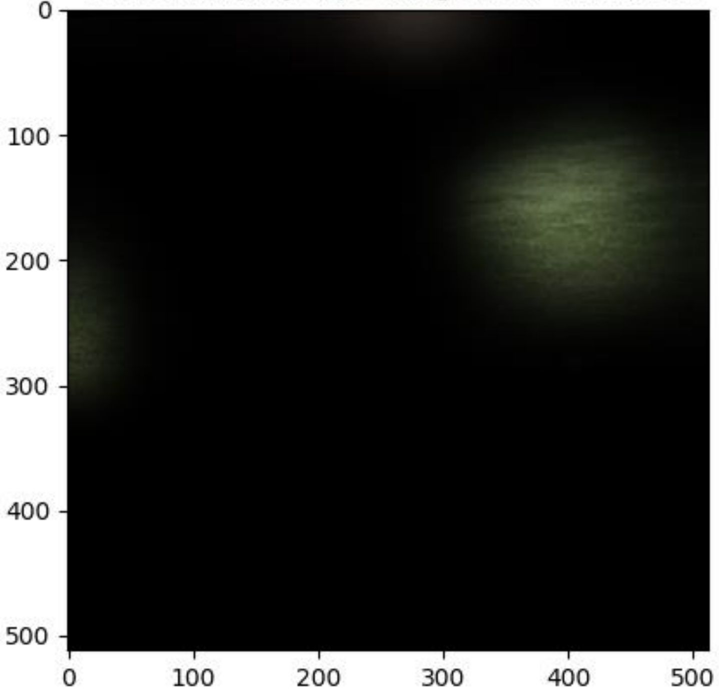
a baseball player pitching **a** ball on a field



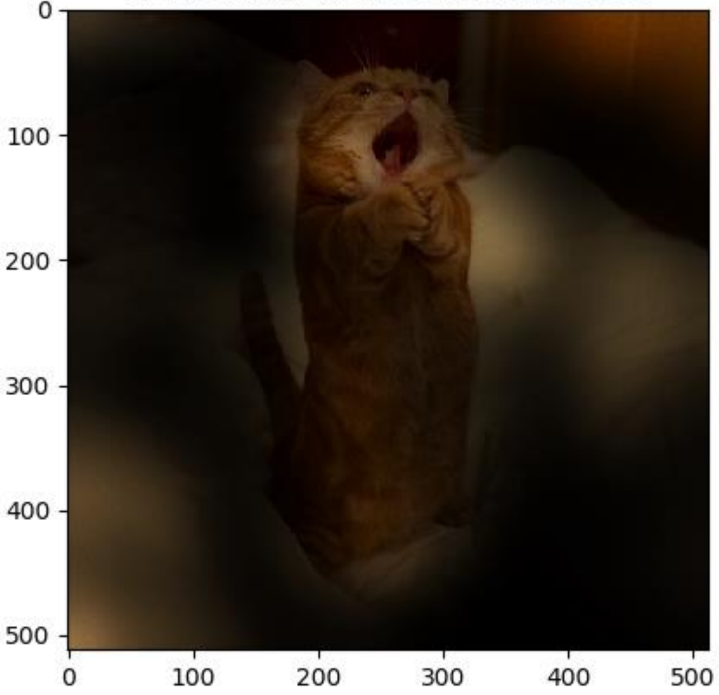
a baseball player pitching a **ball** on a field


a baseball player pitching a ball **on** a field
0
100
200
300
400
500
0
100
200
300
400
500

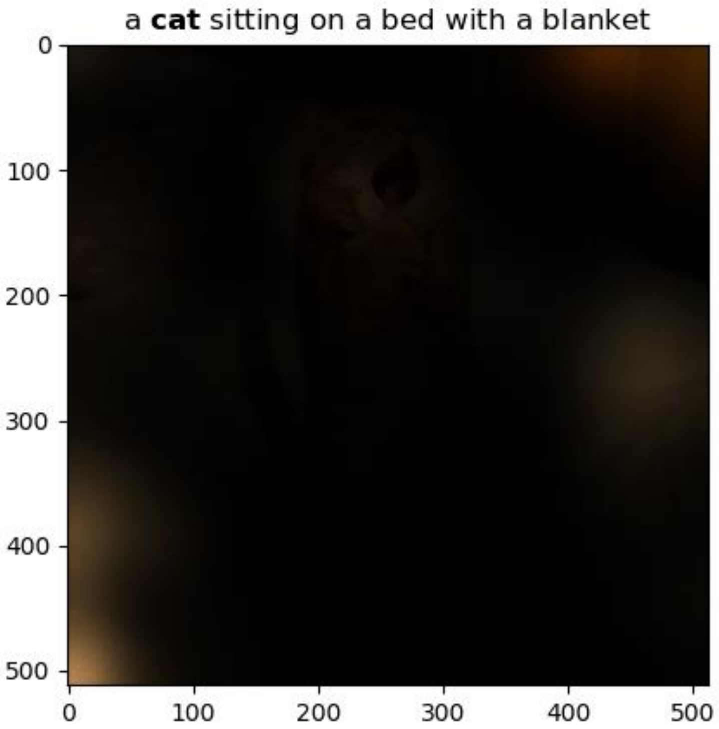

a baseball player pitching a ball on **a** field
0
100
200
300
400
500
0
100
200
300
400
500



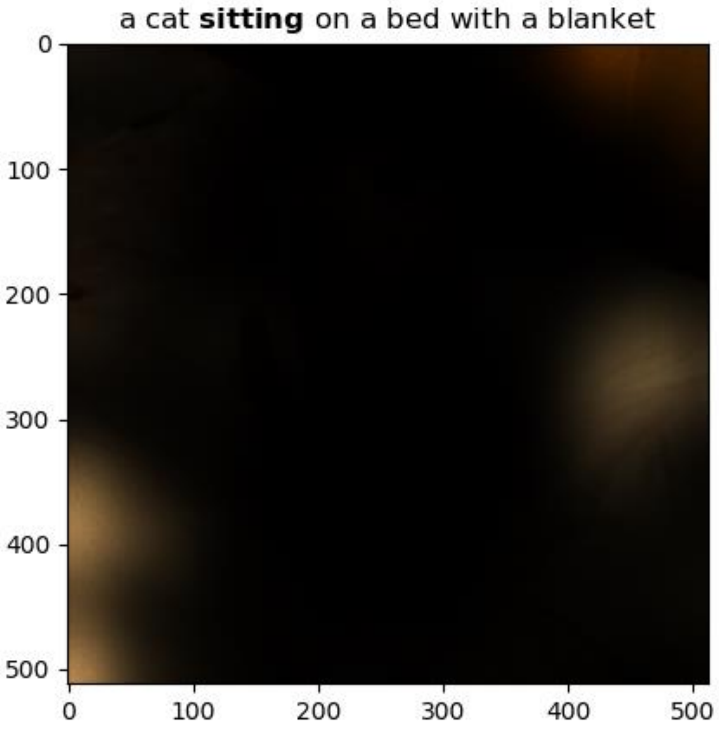
a baseball player pitching a ball on a **field**



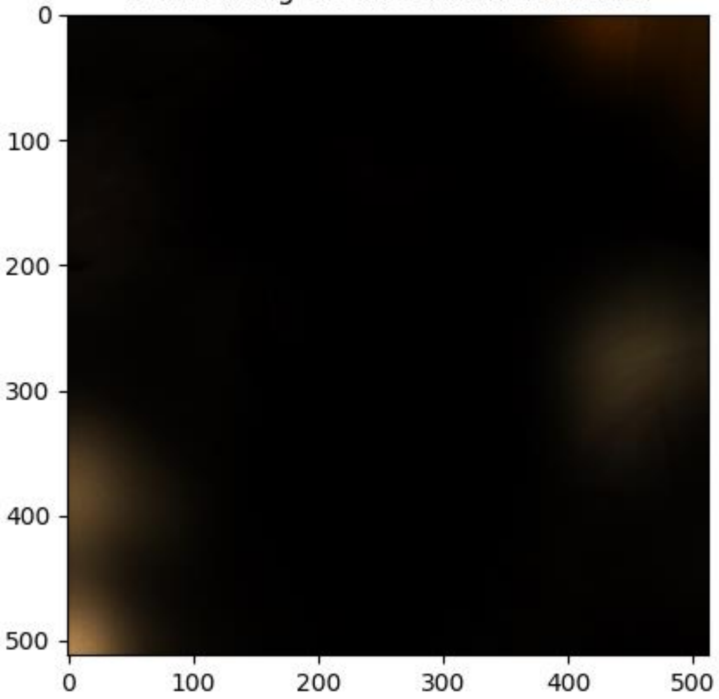
a cat sitting on a bed with a blanket



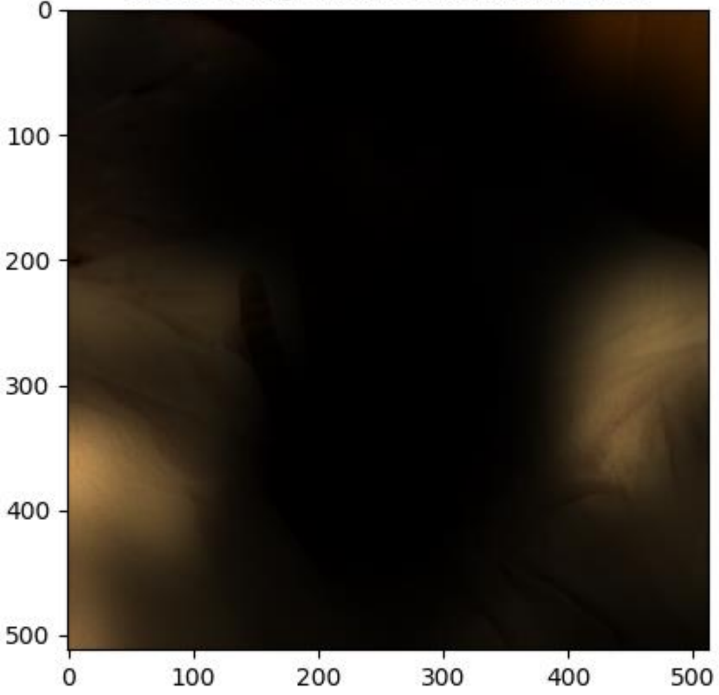
a **cat** sitting on a bed with a blanket

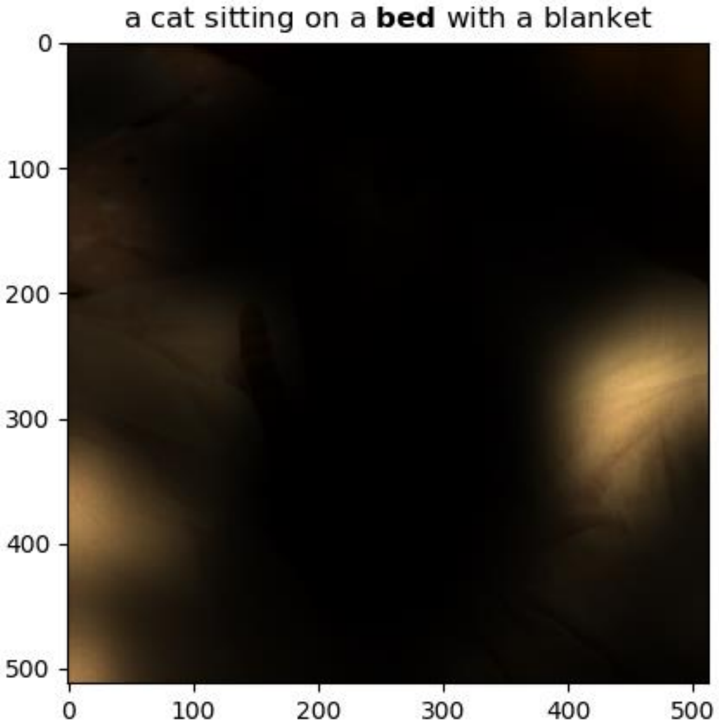


a cat **sitting** on a bed with a blanket a cat sitting **on** a bed with a blanket

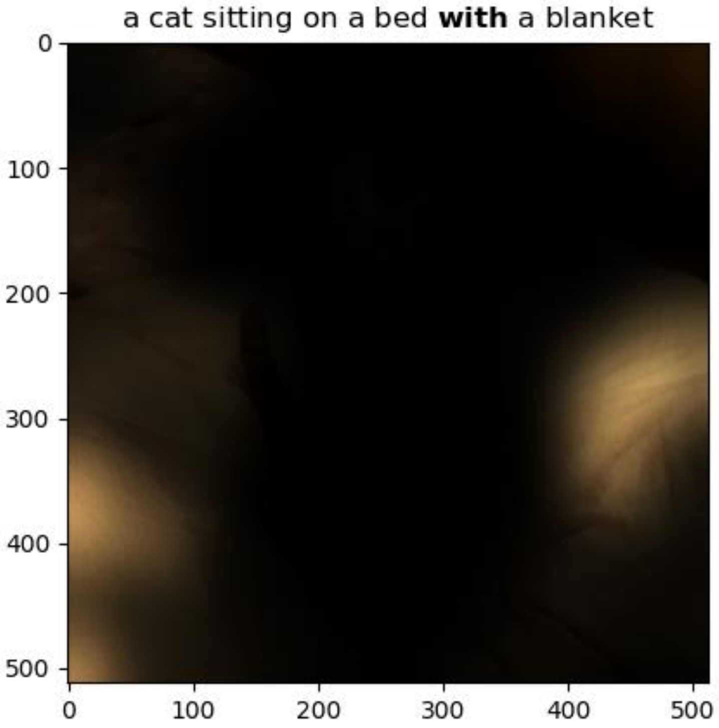


a cat sitting on **a** bed with a blanket

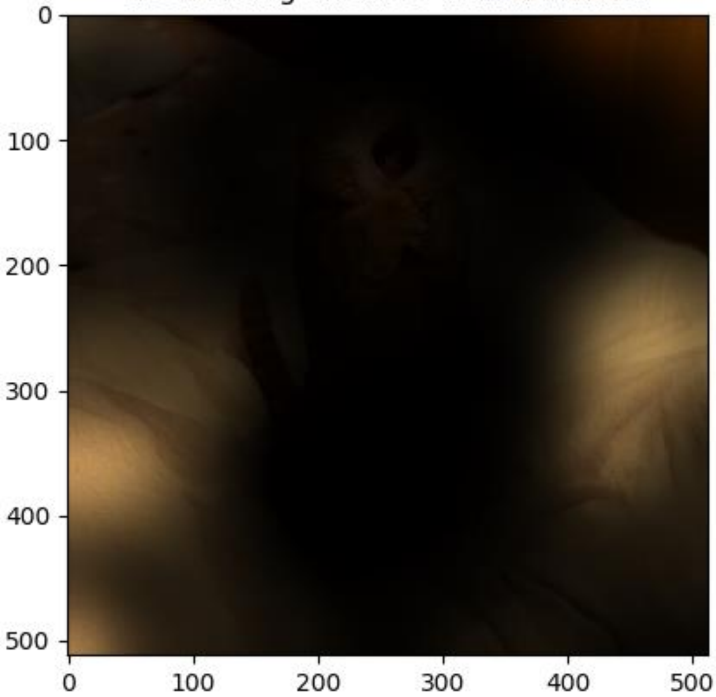




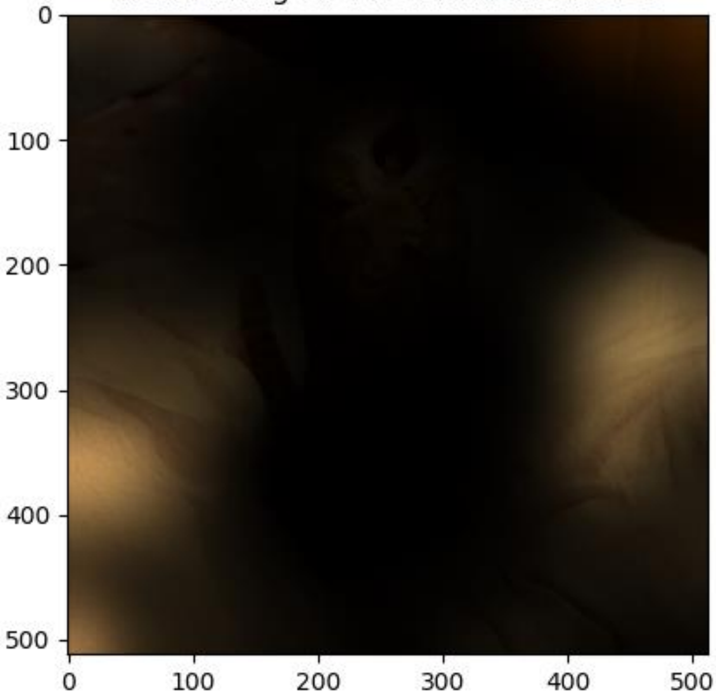
a cat sitting on a **bed** with a blanket



a cat sitting on a bed **with** a blanket



a cat sitting on a bed with **a** blanket



a cat sitting on a bed with a **blanket**